

Group 2

Spesialisering innen forskning i design av IT

What is the impact of AI on people's everyday lives, with a focus on conversational agents

UiO : Universitetet i Oslo



Department of informatics.

Fall 2019

Written by:

Maren Søby Fosser, Tina Steinstø og Suresh Sapkota

Table of contents

1. About us	3
2. Area of interest	3
3. Background	4
4. Questions	5
5. Methods	6
5.1 Literature study	6
5.2 Interviews	6
6. Findings	7
6.1 Interviews	7
6.2 Literature study	9
7. Discussion	11
7.1 How does artificial intelligence impact people's everyday lives?	11
7.2 How does people's expectations of AI affect how they interact with AI?	12
8. Conclusion	14
9. Lessons learned	15
10. References	16
Appendix 1	19
Appendix 2	20
Chatbot assignment	20
Appendix 3	21
Appendix 4	22

Report - iteration 3

1. About us

Our group consists of three members, who are doing masters in Design. Suresh Sapkota , Maren Søyby Fosser and Tina Steinstå. Suresh and Maren are just new to master and started this autumn, and Tina is writing her master thesis already about NAV's use of chatbots. We have all done our bachelors here at Ifi.

2. Area of interest

What is the impact of AI on people's everyday lives, with a focus on conversational agents

We want to research how artificial intelligence impacts people's lives. We have a lot of products on the market these days that claim to have AI, for example virtual assistants or robotic lawn mowers. We find this area interesting because AI technologies gives new opportunities for different kinds of interaction and user experiences.

Because of the scope of the assignment we have chosen to focus on conversational agents, as most people have one pre-installed on their phones. Our question is whether they make people's lives easier or not, as often advertised, or if they create new problems we have to overcome. Users experience with human-AI interaction changes over time Buchner et al. (2013). This will also be an interesting part of our study, as we also want to look at how the expectations people have to the conversational agent affect how they use it and whether they feel that it makes things easier.

3. Background

Verne and Bratteteig (2018) wonders in their article whether AI will make PD (participatory design) obsolete. They conclude that PD is not obsolete because of AI, but that AI poses challenges to a PD process because it is a technology that changes unpredictably over time (Verne & Bratteteig, 2018). We find this interesting in connection to our area of interest because if AI changes in an unpredictable way in a design process, then it stands to reason it does this when it is in use too. An ever changing technology that will continue to develop

as it gets more information about its user might get new roles in how it affects people's daily lives and its role in these.

There has not been done much research on conversational agent used for the purposes that we would like to explore. There has been done other studies, however, like conversational agents that work as a museum guide. A paper from 2005 describes how the CA (conversational agent) 'Max' guides visitors at a museum, in addition to participating in small talk (Kopp, 2005). Another study shows problems that arises with a CA that were designed to give information about train tables in The Netherland (Sturm,1999). But how these CA's work in everyday settings outside of these restricted places and specific contexts has not been a focus in HCI-research, and we would therefore like to explore this further.

We would like to focus on conversation agents used in people's home for everyday tasks and examine the use of personal assistants like Google Home and Siri. Apple's voice assistant Siri is designed to help within specific domains and listens to keywords for identifying the subject area. It can access information from different apps and consider whether it has enough information to complete the task. All inquiries are processed through Apple's servers and this gives Apple a huge amount of data that can be used to improve the service. (Aron, 2011) Google Home works a bit differently as it is a separate device that is activated when you use specific keywords. It can play music and access video content and can be the control centre for your home. (Tillman, 2019)

Conversational agents are relying more and more on automation. An issue with automation is the type of feedback that is provided and whether the design is made for the existence of errors. Normans article from 1990 explains how problematic inappropriate feedback and over-automation can be. It is not the automation in itself that is the problem, but how it is designed. Feedback is important because equipment can fail and unexpected situations can arise. (Norman, 1990) This can also be linked to the design of conversational agents. There can be errors and unexpected situations with this type of technology too, and the type of feedback the conversational agent provides is therefore important. Too much automation can have consequences as the more decisions you let your CA make, the more control it has over your life. As these decisions are made automatically, often without notifying the user, it can be difficult to know which decisions are being made on your behalf and the implications it could have.

4. Questions

1) *How does conversational agents impact people's everyday lives?*

With this question we would like to understand in what way conversational agents is incorporated in people's lives. Whether it helps them do tasks or for other purposes, and how this is done. We would also like to know how conversational agents has changed the way these tasks were done before. In addition, it is relevant to understand if conversational agents still has a future in people's life, and how its role will evolve.

Some robots require additional work done by the users, like for example the vacuum cleaner Roomba that requires that the user tidy up and make it easy for the robot to move around the room (Forlizzi, 2006). How do these additional tasks affect how the user interacts and how does tasks change because of the use of conversational agents and AI?

To seek information about this question, we would like to talk to users of conversational agents by doing interviews.

2) *How does people's expectations of AI affect how they interact with AI?*

There has been conducted several studies about how user expectations can affect how the users interact. Often there is a gap between user expectation and system operation because of the users mental model and lack of feedback that can affect the interaction (Luger & Sellen, 2016). Some users will have too high expectations of what AI can do or expect something different from what the AI has been programmed to. How does this affect the use and perception of conversational agents?

According to Hectht (2018), in developing new technologies it is important to take the public's view into consideration. We would like to understand what the expectations were of the people we interview and if these were met or not.

5. Methods

5.1 Literature study

We would like to do a literature study because research has been done on these subjects before. We want to know what topics have been researched so far and what we would be interesting to investigate further. The literature review will be the basis for our interview guides and we will also use findings from our literature review to analyse the findings from the interviews. We will be reading articles that concern both the expectations of AI and thoughts about the future of AI. We think these factors will affect how the conversational agents are used and to what extent. Because some of what we want to know is speculation, we think it is important to see what other researchers predict about the future of AI.

To analyse our findings about the users expectations we will read 'Like having a really bad PA' and see how our findings from our interviews align with the research done on this topic. We will read articles about the use of conversational agents to learn more about what have been researched previously, but as most of the research on conversational agents have been done in constricted settings within specific contexts, we will use interviews to explore the impact of these agents on people's personal life. We will also use the guidelines from Amershi et. al (2019) to reflect on whether the CA's of our interview objects fulfill their needs and expectations in relation to the user experience they have with their CA's. We find the guidelines useful because they might explain the bad experiences, as well as the good ones of our interview objects.

5.2 Interviews

We have chosen to conduct interviews to get new knowledge about how people use personal assistant technologies, and what their experience in using them is. We have chosen to do interviews because we think it can give interesting and new knowledge when compared to studies other researchers have done.

Because several of the research papers in this field is written in English speaking countries, we think it is also interesting to acquire knowledge from people that have a different cultural backgrounds and use these assistants in languages that are not the mother tongue of the developers behind the technology.

To execute the interviews we have made an interview guide with some questions we think will be able to give us valuable information. The guide has some questions that are defined, but we aim to ask the interview objects to elaborate on their answers by asking follow-up questions. We will in other words conduct semi-structured interviews.

6. Findings

6. 1 Interviews

6. 1. 1 Interview objects

We have chosen to interview people who use personal assistant technology. The reason behind this decision is that there is a lot of AI technology out there, and we need to narrow down our area of interest to be able to present interesting findings. We think people who use conversational agent technologies can give us valuable insight into how AI affect their lives, and it is also a relatively new technology that is interesting in regards to how people interact with AI.

We recruited three interview objects by contacting acquaintances we knew had smartphones that include a conversational agent or owned a home assistant like Google Home. Because the interview objects knew us, we have to take into consideration that some of the answers might be affected, as well as our own bias. Nevertheless, we argue that their insight is valid in this setting as we are not asking questions that are sensitive or embarrassing.

Our interviewees were in different age groups, from 24-65, and had different backgrounds when it came to technology. The younger participants studied technology, but belonged to different areas of research, and the older participants had less of a technical background.

6. 1. 2 Use

Our interviewees had different experiences with AI and the interviews we conducted gave us therefore insight into different ways of use. The people we interviewed had used Google Home, Google's assistant on phone and Siri. They used them for different purposes like setting alarms and searching for things, like they do in a 'traditional' Google search or finding directions. One of the interviewees stated that she used it more for fun than for

specific purposes. None of our interviewees used their personal assistant for very meaningful or important tasks.

6. 1. 3 Expectations

The students in technology searched for what Siri or Google Assistant could do and when new functions became available they started exploring. They had not thought about all the functions the technology could have and there seemed to be several functions they did not know about.

When the students started using personal assistants they thought the assistant would be more capable of doing things than they were. The older interviewee started using the personal assistant out of curiosity and to explore the new technology.

6. 1. 4 Benefits

The participants gave us different answers when it came to whether or not the personal assistant was effective in use for them. The older participant experienced that often it would be faster to type than talk, because the assistant misunderstood her a lot. One of the younger interviewees on the other hand thought the opposite.

6. 1. 5 Challenges

All the interviewees had problems being understood by their assistants, but the female, older interviewee more so than the males. She had little trust in Siri when it came to being able to do tasks efficiently.

They all experienced that words they said were misunderstood by their personal assistants, and that they would get wrong results based on this. In some cases it was also necessary to be very specific, but not in an intuitive way. The older participant explained, for example, that to check the weather in Haukeli she had to ask about Haukeli Fjell (Haukeli mountain) to be able to get the right results.

One of the interviewees had problems with false positives, where his Google assistant would start in the classroom when it thought someone said "Hey Google". He also experienced that when he actually wanted to start a dialogue that Google did not respond every time, and required more than one try.

One of the students said that they liked using the assistant in the beginning, but could never get used to talking to a phone especially in public.

6. 1. 6 Future use

All the interviewees stated they wanted to continue using their personal assistants, and expressed a prediction in that the technology will become more useful in the future. One of the younger participants said he thought that with Internet of Things, personal assistants would probably be a lot more integrated in our daily lives than at the moment.

6. 1. 7 Impact on everyday life

Our interviewees had not used conversational agents regularly over a longer time period and the tasks they used it for did not impact their daily routines. It was therefore difficult to see specific examples where it had impacted their lives. When we recruited our interviewees we had hoped to find someone who used conversational agents regularly, but this proved to be difficult. Most of the people we know that use conversational agent use them occasionally, not regularly. However, all the users thought that in the future they would use their conversational agents more which could lead to a bigger impact on their lives in the future. The reason they said this was because they thought the technology would improve over time.

6. 2 Literature study

6.2.1 How does Conversational agents impact people's everyday lives?

A survey for Google Home from 2019 shows that the most common tasks Google Home is used for, is listening to music and asking about the weather.(Sterling, 2019) This indicates that conversational agents are mainly used for specific tasks, and the users rarely explore the other functions you could use the system for.

It can be easy to forget that a task can be performed by simply speaking out loud, as there are no visual clues to remind the users of the voice command function. Because of this there has been discussions about whether touch screen would make it easier to use voice command based devices like Google Home.(Condliffe, 2017) With a screen you could easier to explore the device and you could receive unobtrusive notifications.

Current conversational agents have many challenges in maintaining a sustainable human-AI interaction. Følstad & Brandtzæg (2017) describes in their article that technology giants like

Google, Facebook, and Microsoft claim their users will be moving their digital interaction from websites and apps with graphical user interfaces to natural language interface platforms such as Messenger and Allo. In the future, chatbots may be the preferred user interface for many activities like communication. This will probably also be the case of other types of conversational agents, for example Siri or Google Home. If this happens, huge challenges and opportunities await in the field of HCI. Følstad & Brandtzæg (2017) claims these implications for HCI can be important in the rapid development of chatbots and on how we approach the design of them. To approach these challenges it is important that developers have to move from explanatory task to interpretive task. Among other things, make it visible to the user what features are available in the system and addressing users need (Følstad & Brandtzæg, 2017). Følstad & Brandtzæg (2017) proposes that to solve the current issue of chatbots, HCI need to reorient new practices and research to meet the challenges of chatbots and natural language interfaces. The success of chatbots and other conversational agents depends on the ability to hold a conversation thread in multiple stages without the breakdown of conversation (Følstad & Brandtzæg, 2017).

Buchner et al. (2013) report how adoption processes of robots in factories changes over time. It takes time for users to understand the functionality. When a user interacts with AI for a longer period of time, their experience regarding the AI changes. Most people find it easier to understand the system after continuous interaction whereas some may find it difficult to start due to its advanced level of user interface (Buchner et. al., 2013). This can also be used to understand the learning and initial adoption processes of conversational agents.

To create better user experiences in AI systems, Amershi et. al. (2019) present guidelines that are meant to support this. We think reflecting and comparing the 18 guidelines with our interviews will give a better understanding on the impact of AI on people's everyday lives because it might reveal the reason behind positive or negative experiences.

6.2.2 How does people's expectations of AI affect how they interact with AI?

User expectations can be a source of many challenges. A study from 2019 (Luger & Sellen, 2019) shows how the users mental model will affect how they expect the system to react and what they expect the system to be capable of. Without knowledge about the limits of the system and how it works it is difficult to make accurate judgments about a systems capability. When a system did not do what the user thought it should be capable of doing

the user would end up abandoning the task. Lack of feedback regarding how the system worked was therefore an issue. The users with more knowledge of computer science, however, would try again in a different way. As the users learned to use their conversational agent they relied more on keywords and shortened their sentences because they understood more of how the system interpreted what they said. Bridging the gap between the users mental model and the system's capabilities is therefore important (Luger & Sellen, 2019). To bridge this gap system feedback and how the CA is designed, is important. We would like to further explore how feedback is interpreted based on the mental model the user have and how this can be improved for a better user experience. In our study, we will therefore focus on exploring further how expectations affect the interaction and uses of the CA.

Another article that shows how the user's initial expectations affect the final perception of a system is 'Will You Accept an Imperfect AI?'. This study shows among other things how things like false positives and false negatives is interpreted by the user. Too many false positives can lead to unwanted requests leading to frustration and too many false negatives can lead to the AI missing requests. Another important fact shown in this study is that if the user contributes to the system they are also more accepting of mistakes (Kocielnik et. al., 2019).

7. Discussion

7.1 How does artificial intelligence impact people's everyday lives?

7.1.1 Violation of guidelines

We used the guidelines provided by Amershi et. al. (2019) as a tool to understand where our interviewees might have had problems with their conversational agents. We found that several guidelines were violated for some or all of the interviewees. All of the interviewees had problems with being understood because of their accents, which is a breach with guideline G5, that the system should deliver expected results in the user's social or cultural context (Amershi et. al., 2019). One of these interviewees was female, so we cannot be sure if it was her accent or the fact that she was a woman that the conversational agent did not understand her. If it was the latter, it is a breach of guideline G6, that the system should mitigate biases (Amershi et. al., 2019).

All the participants expressed that they were not aware of all the functions their conversational agents had. This is a breach of guideline G1, which states that the system should make clear what it can do (Amershi et. al., 2019).

We found that the AI has made both positive and negative impact in people's everyday life. One interviewee felt using their conversational agent was more time consuming to use than to open apps and search for information manually. In contrast another interviewee felt using theirs was quicker than typing. We found that all of our interviewees think their conversational agents need improvement, and from the guidelines presented in the article by Amershi et. al (2019) we can get insight into what improvements can be made.

7.1.2 Understanding functionality

We found that none of our interviewees knew about all or most of the functionalities their conversational agents had to offer. This is not an uncommon occurrence according to Condliffe (2017). Condliffe (2017) proposes in their article that a touch screen might be available with home assistants like Alexa or Google Home in the future, to support further use for the owners of the products. This might make it easier for users to understand the capabilities of the system, and the system could offer more guidance when being in use.

7.1.3 (Dis)Continued use

Users of technologies in factories need time and continuous use to understand the extent of the functionality (Buchner et. al., 2013). We find this transferable to the experience our interviewees had with their conversational agents, as they generally used little time on them. Our interviewees found their conversational agents interesting in the beginning, but we found they had problems discovering new functionalities after the initial interest. The reason behind this might be because they did not take the time to learn them.

7.2 How does people's expectations of AI affect how they interact with AI?

7.2.1 Expectations leading to disappointment

Kocielnik et. al. (2019) did a study on a scheduling assistant where they found that user's expectations were related to the satisfaction of the system. They claim that very high expectations negatively affected perception and acceptance of such systems (Kocielnik et. al., 2019). Through our own interviews we saw that our interviewees had high expectations of their CA's in the beginning, and were intrigued because they thought it was very smart. In the end none of them used their CA's that much. We understood this to be because they

did not find them very useful and because they found difficulties. Among these were not being recognized because of their accents or false positives where the CA initiated a conversation at inappropriate times.

Bhattacharjee (2001) present a model called Expectations Confirmation Model in a study from 2001. This model showed that the satisfaction of the user primarily was determined by the user's confirmation of expectation, and that confirmation had a significant influence on post-acceptance (Bhattacharjee, 2001). It might be the case that our own interviewees had very high and unrealistic expectations of their CA's and were put off by them when they saw the system could not live up to these.

7. 2. 2 Understanding the system.

As AI is gradually developing, human expectations on AI is also increasing (Grudin, 2009). We found in our interviews that our interviewees had high expectations of what their conversational agent could do, but were disappointed with the limitations when they actually used them.

According to Endsley (2011), the use of situation awareness is key to understanding decision making and performance in human-AI interaction. Developing design methods with situational awareness involve giving users more control in different situations, and can contribute to good user-centered design (Endsley, 2011). When a system lacks situational awareness a user might have problems understanding exactly what the system does (Endsley, 2011). Our interviewees mentioned not being aware of all the functions their conversational agents had, as well as not understanding how the system did what it did. These are factors that might have affected their user experience and given poor situational awareness.

7. 2. 3 Bridging the gap between user expectations and the capabilities of the system

Our findings show how important it is to bridge the gap between user expectations and experiences. A system will often convey limitations if it fails to do a task, however it will often not make its limitations and capabilities clear to the user when this is not the case. (Luger & Sellen, 2019) This affect the expectations the users have which we also saw in the interviews we conducted. When they expected it to be able to do something it could not do, they were disappointed by its limits, especially as they were not aware of the limits beforehand. There were also functions they were not aware of as they did not know the

capabilities of the system. To bridge this gap we think it is important to think of the implications when designing features like feedback. If for example Siri had notified the user that place names needed to be stated completely, getting weather report for correct location would not have been an issue. If all the functions of the personal assistant had been clear beforehand, the user would have known more about the potential of the CA. Bridging this gap would therefore probably mean both less disappointment and more use of the CA for different purposes.

8. Conclusion

Our findings from our interviews show that the users had several problems with their conversational agents. They did not know all the tasks their conversational agent could help with and the CA did not make it clear what it could do and what went wrong when it could not complete a task. We saw that several of these difficulties was because of violation of guidelines and that the interface made it difficult to know what the CA could do and how it worked. The CA's did not manage to uphold guidelines like 'Make it clear what the system can do' and 'Make it clear how well the system can do what it can do'(Amershi, 2019) which lead to several disappointments.

The users we talked to refrained from using their CA's regularly because their expectations did not match what the CA was capable of. Too high expectations made them them disappointed when it could not complete a task they thought it should be able to and made it less tempting to try the CA for similar tasks. When it could not complete a task they thought it should, they assumed that it could not do other things as well.

We can see from our literature study that AI have the ability to make significant impacts on people's life and work (Norman, 1990), but because the CA were not used regularly this was not the case in our findings. However, had the CA been able to do what the users expected it to, the pattern of use might be different. Had the CA made it clear what it could and could not do, the users mental models would align with the system capabilities and they would probably have been more in use. CA have the possibility of affecting people's life because of all the data that is gathered and the decisions that are made based on this information on behalf of the users. We can therefore see from our findings that more realistic expectations could lead to a use of CA that could have an impact on people's life. But as the use of CA today, our findings show that use of conversational agents does not have a visible impact on the user's everyday life.

9. Lessons learned

During this project we learned several things about researching use of AI and CA. We saw that even though many AI systems are used in our everyday life, conversational agents are not used as regularly yet.

We realised that researching the impact of AI in everyday life was more difficult than we had thought originally. There could be two different reasons for this. Either it is because of the participants we recruited, which might not be as representative for the user group as we thought, or it could be because AI does not have as big an impact as we thought. We tried recruiting people with different ages and different technical background to make sure of the reliability of our study, but as we could not find much research done on the use of personal assistants in everyday life, it was difficult to know who would be the best to talk to and what kind of use would most likely impact someone's life.

However, what we learned about expectations in the literature study aligned well with what we found in our interviews. It was therefore easier to see specific examples of how expectations affected use. On the other hand, many of the studies done on CA were in specific contexts and not for everyday use. This therefore seems like a topic that would have been interesting to explore further.

10. References

Aron, J. (2011) 'How innovative is Apple's new voice assistant, Siri?' *New Scientist* 29. oct. 2011. Retrieved from:

<https://www.newscientist.com/article/mg21228365-300-how-innovative-is-apples-new-voice-assistant-siri/>

Amershi, S., Weld, D., Vorvoreanu, M., Fournery, A., Nushi, B., Collisson, P., ... & Teevan, J. (2019). Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM.

Bhattacharjee, A. (2001). Understanding Information Systems Continuance: An Expectation-Confirmation Model. *MIS Quarterly*, 25(3), 351-370.

Buchner R., Wurhofer D., Weiss A., Tscheligi M. (2013) Robots in Time: How User Experience in Human-Robot Interaction Changes over Time. In: Herrmann G.,

Condliffe, J. (2017) 'AI Voice Assistant Apps are Proliferating, but People Don't Use Them' *MIT Technology Review* 23. jan 2017. Retrieved from:

<https://www.technologyreview.com/s/603420/ai-voice-assistant-apps-are-proliferating-but-people-dont-use-them/>

Endsley, Mica R.. Designing for Situation Awareness: An Approach to User-Centered Design, Second Edition CRC Press. 2011 (chapters 2 and 10)

Forlizzi, J. & DiSalvo, C.: Service Robots in the Domestic Environment: A Study of the Roomba Vacuum in the Home, 2006.

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. interactions, 24(4), 38-42.

Grudin, J. AI and HCI: Two Fields Divided by a Common Focus. *AI magazine* 30, no 4 (September 18, 2009).

Hecht, J. (29.09.19) Managing expectations of artificial intelligence *Nature*

<https://www.nature.com/articles/d41586-018-07504-9>

Norman, D (1990). The problem of automation: Inappropriate feedback and interaction, not over-automation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, Vol. 327, No. 1241, *Human Factors in Hazardous Situations* (Apr. 12, 1990), pp. 585-593

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019). Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (paper no. 411). ACM.

Kopp S., Gesellensetter L., Krämer N.C., Wachsmuth I. (2005) A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application. In: Panayiotopoulos T., Gratch J., Aylett R., Ballin D., Olivier P., Rist T. (eds) *Intelligent Virtual Agents. IVA 2005. Lecture Notes in Computer Science*, vol 3661. Springer, Berlin, Heidelberg

Luger, E., & Sellen, A. (2016, May). Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*

Sterling, G. (2019) Search Engine Land, 21. march 2019. Retrieved from:

<https://searchengineland.com/survey-finds-google-home-users-do-more-have-far-higher-satisfaction-than-alexa-owners-314379>

Sturm, J., Os, E., Boves, L. (1999). Issues in Spoken Dialogue Systems: Experiences with the Dutch ARISE System. In the proceedings of ESCA Workshop on Interactive Dialogue in Multi-Modal Systems (UDS-99), Kolster Irsee, Germany.

Tillman, M. and O'Boyle, B. (2019) 'What are Google Home, Home Max, Nest Mini, Nest Hub and Nest Hub Max and what can they do?' *Pocket lint* 15. oct. 2019. Retrieved from: <https://www.pocket-lint.com/smart-home/news/google/137665-what-is-google-home-what-can-it-do-and-how-does-it-work>

Verne, G., Bratteteig, T., 2018, Does AI make PD obsolete?; exploring challenges from Artificial Intelligence to Participatory design

Appendix 1

Interview guide

What kind of AI assistant do you use?

How long have you been using your personal assistant?

How often do you use your personal assistant?

Have you used similar technologies before?

What were your initial expectations of your personal assistant?

What were you planning to use it for? Do you use it for this purpose now?

Can you tell me something about your experience in using this assistant?

-any bad experience you have regretted using AI?

What do you use AI for? (tasks?)

Do you think <insert AI name here> does this task efficiently?

Do you want to keep using this AI?

How do you think this personal assistant will evolve in the next years? (What new functions/tasks will it be able to do?)

Appendix 2

Chatbot module 1

Chatbot assignment

About the process?

We started by finding an idea, and we chose to imagine we would make a chatbot for a company, Visit Oslo. We used their website as inspiration for what a user might want to know more about, and take inspiration from their information flow.

We wanted to make sure the chatbot will make continuous conversational threads during the interaction with the user. The successful chatbots, have the ability to hold conversation thread in multiple stages without breakdown of conversation (Følstad & Brandtzæg, 2017).

We had multiple iterations, to figure out if the chatbot actually complete the task without failing during the interaction with users. We tested the chatbot and went back to make some changes before we tested the chatbot again.

What did we learn?

We learnt that designing a conversation requires a lot of effort, and we had to make very different decisions on how the user interacted with our bot in comparison to how we normally design user interactions. The bot was not as smart as we thought with the keywords, so we had to write more than we initially thought necessary.

Because of the type of interaction, it was difficult to predict what a user might think faced with the questions the chatbot asked. We saw, chatting with our bot, that we forgot what we expected the user to respond, and noticed some questions would probably seem very ambiguous for a first-time user. Several times during the interaction, our bot failed to connect or understand what the user is looking for, we thought it was because of input we had provided to our bot. We tried instead to make the bot come with suggestions on what kind of input the user might ask, to nudge them in the right direction or give inspiration.

Appendix 3

Chatbot module 2

Chatbot design task : Reflections about making changes to the program

The second task in module 2 was to learn more deeper on how AI and chatbots work. The internal functionality of chatbot, input provided to it and changing the different aspects of the code.

At first we tried to read through the code to understand it. We found it a bit difficult to understand every aspect of the code. When we tried to make changes, we got an error message, but we realised after a while that we had forgotten to upload the file. As we had not used the program before so it took some time to figure out how to fix this error.

When we understood how to make changes, we started changing the number of maximum words, which show no changes in the accuracy and we increased the number of epochs to 3, and again no changes in accuracy since the accuracy already reached its maximum after epochs.

Then we starting changing dense layer. At first we had some issues when we tried to change this such as; the program could not compile at all we got no results. The result of changing max words was that we got more numbers displayed on the screen, and we found out that AI have more accuracy at 8 or above. Below 8 the accuracy started to decrease. We found out that the higher dense value makes AI more effective and maximum potential to memorize the data. We also tried to change the input text and max words.

We struggled a bit with the program and it did not work every time we tried to run the code. Even though we had some challenges while implementing changes it was interesting to see how the changes we tried to make affected the program, but it would probably be easier to make changes if we knew more about how the different parts of the code worked.

Appendix 4

Module 3: Evaluation of Netflix

Subject and the scope of the evaluation

We have decided to evaluate Netflix's recommendation system. Netflix uses artificial intelligence to recommend series and genres based on the series you have recently watched or rated.

We will evaluate the issues with Netflix. How long time we have spent in evaluating. We have selected this system because we have been using this for a long time period and this will make it easier to evaluate the AI behaviour in Netflix, as we have experience with the system over time. It is also an interesting system as the recommendations that are personally made for you appear on your profile without notifying you of how the system works and recommends things for you. Because of this we thought that it would be interesting to see how this system works according to the guidelines.

Evaluation plan

1. Browse website

By browsing the website we can evaluate and find (or not find) the functions the Netflix recommendation system has to offer.

Guidelines we evaluate doing this activity:

- G1: Is it clear what the system can do?
- G2: Make clear how well the system can do what it can do.
- G5: Match relevant social norms.
- G8: Support efficient dismissal.
- G9: Support efficient correction.
- G10: Scope services when in doubt.
- G11: Make clear why the system did what it did
- G12: Remember recent interactions
- G17: Provide global controls.
- G18: Notify users about changes.

2. Try search engine

By using the search engine we can see what Netflix would recommend us to watch outside of the front page categories.

Guidelines we evaluate doing this activity:

- G4: Show contextually relevant information.

3. Give high ratings to films you do not like

We want to see how our recommendations change if we indicate we like films we normally would never watch.

Guidelines we evaluate doing this activity:

- G6: Mitigate social biases
- G11: Make clear why the system did what it did
- G13: Learn from user behavior.
- G14: Update and adapt cautiously
- G15: Encourage granular feedback.
- G16: Convey the consequences of user actions.

Evaluation results

We focused on two categories during our evaluation process. 1) We tried to find whether the guidelines were relevant in the system we evaluated. 2) Is it easier to understand the guidelines. Our result is based on these two questions.

Browse website

Guideline		Verdict
G1	Is it clear what the system can do?	This seems to be not clearly specified by the system. But as we have experienced, it is clear that you can search series or genres in search engine.

G2	Make clear how well the system can do what it can do.	With each recommendation the system states a percentage of how relevant this match is for you. However, if the percentage is too low the system will not give a match-percentage.
G5	Match relevant social norms.	There is an option for parental control on Netflix accounts that limits what kinds of films and series are recommended. This is an account automatically generated called “barn” in our case. This account has a completely different design that is more whimsical than the standard, and it groups up films and series not only by genre, but also famous characters like Barbie or Peppa Pig. It is also possible to put an age restriction on an already existing account, which would maybe be more suitable for pre-teens or older children.
G8	Support efficient dismissal.	This system does not allow you to delete/dismiss the list of recommendations but you can easily scroll down to the next suggestion list.
G9	Support efficient correction.	You can give thumbs up/ down to the series that are recommended to you.
G10	Scope services when in doubt.	When you skip the list of some recommendations made by the system, it automatically changes the list and provides you new list of series based on other users’ ratings/views. Such as list of “popular on Netflix”.
G11	Make clear why the system did what it did	You cannot see why each series is recommended to you. Some series will be recommended based on specific series you have seen, but the categories you have on your profile do not have an explanation to why they appear on your profile.
G12	Remember recent interactions	The system provides the list of words based on the letter/s you type in search engine. And the list of related

		words will pop up as “explore title related to:”
G17	Provide global controls	The system does not let its users turn off the recommendation system. It is only affected by the users history.
G18	Notify users about changes	<p>When changes are being made, you will usually receive a mail or notification of changes made. But these are changes regarding series and films added to Netflix, not how the recommendation system works.</p> <p>Some of us have had Netflix for years (2014), and when the review system was removed in 2018 we did not receive a notification or email about this.</p>

Try search engine

Guideline		Verdict
G4	Show contextually relevant information.	You will receive results based on what is relevant to the genre of what you are searching for, not the specific words in the title. This shows that the system knows which genre a film or series belong to, even if the series you are searching for is not available for you.

Give high ratings to films you do not like

Guideline		Verdict
G6	Mitigate social biases	By rating films that consists of similar themes we

		<p>received more recommendations that fit within these. This can create an echo-chamber where users only watch a certain type of film, although Netflix does not in any way hinder you from trying something new, as some of the recommended columns, like “popular on Netflix”, are based on not only one user, but many.</p> <p>It is possible to claim that a user with very strong opinions on something, for example women’s role in society, can have their views reinforced by only being recommended James Bond-esque films.</p>
G11	Make clear why the system did what it did	<p>It was not always clear why Netflix recommended some of the films. Most of them have a compatibility percentage, and there is no explanation on how this is calculated.</p> <p>However, when we liked some Bollywood films (that we normally do not watch), we got a lot of recommendations on other Bollywood films, and it was not very hard to guess why.</p>
G13	Learn from user behavior.	This feature satisfies to a larger extent. It recommends you the list of possible series you may like based on your recent activities.
G14	Update and adapt cautiously	When updates are being done, this does not interfere with how you currently use the system.
G15	Encourage granular feedback.	You can use the thumbs up/ down function. When we used this function we could see that the series recommended for us, changes as we gave thumbs up to series we normally would not watch. But apart from this, it is difficult to give feedback to the recommendations you receive.

G16	Convey the consequences of user actions.	Users are not made aware of how the interaction with the site will affect the recommendations that you receive. It is unclear whether the thumbs up/down function and the series/films you watch are being weighted differently when it comes to the recommendations you receive, but we could see that the users actions had consequences for what types of series you were recommended. Our recommendations changed based on how we gave thumbs up/down, but we did not get any notifications of how this action would affect the recommendations, even though we could see that it had an impact.
-----	------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

What we learned by using the guidelines

We learnt that it was surprisingly difficult to teach the recommendation system what we wanted to show up for us without having to actually watch something. It required very many thumbs-ups. We do not know if this is because these compete with all the films and series we have watched over the years or not.

The system is quite subtle in some ways, as one of our group members did not know that most of the categories are tailor-made to individual users. It does not give the user much opportunity to give feedback or correct what is being recommended. There is also little information about how the recommendation system works, even in the help and support pages of Netflix. We also found that the guidelines were not easily understandable and hence it made user a bit difficult to analyze the system using the guidelines.